



融合金字塔和多尺度注意力的多曝光图像融合优化算法

王卫苹, 刁亚鹏

引用本文:

王卫苹, 刁亚鹏. 融合金字塔和多尺度注意力的多曝光图像融合优化算法[J]. *智能系统学报*, 2025, 20(4): 916-927.
WANG Weiping, DIAO Yapeng. Optimized multi-exposure image fusion algorithm integrating pyramid and multi-scale attention[J]. *CAAI Transactions on Intelligent Systems*, 2025, 20(4): 916-927.

在线阅读 View online: <https://dx.doi.org/10.11992/tis.202406032>

您可能感兴趣的其他文章

双向特征融合与注意力机制结合的目标检测

Target detection based on bidirectional feature fusion and an attention mechanism
智能系统学报. 2021, 16(6): 1098-1105 <https://dx.doi.org/10.11992/tis.202012029>

利用残差密集网络的运动模糊复原方法

Image restoration with residual dense network
智能系统学报. 2021, 16(3): 442-448 <https://dx.doi.org/10.11992/tis.201912002>

基于注意力机制的显著性目标检测方法

Salient object detection method based on the attention mechanism
智能系统学报. 2020, 15(5): 956-963 <https://dx.doi.org/10.11992/tis.201903001>

基于注意力融合的图像描述生成方法

An image caption generation method based on attention fusion
智能系统学报. 2020, 15(4): 740-749 <https://dx.doi.org/10.11992/tis.201910039>

高斯核函数卷积神经网络跟踪算法

Convolutional neural network tracking algorithm accelerated by Gaussian kernel function
智能系统学报. 2018, 13(3): 388-394 <https://dx.doi.org/10.11992/tis.201612040>

行人重识别研究综述

Survey on pedestrian re-identification research
智能系统学报. 2017, 12(6): 770-780 <https://dx.doi.org/10.11992/tis.201706084>

DOI: 10.11992/tis.202406032

网络出版地址: <https://link.cnki.net/urlid/23.1538.TP.20250221.1156.010>

融合金字塔和多尺度注意力的多曝光图像融合优化算法

王卫苹, 刁亚鹏

(北京科技大学 计算机与通信工程学院, 北京 100083)

摘要: 为解决复杂光照条件下真实场景中导致的图像噪声、模糊和细节丢失问题, 本文提出一种多曝光图像融合技术 DPEPA-MEF (deep pyramid exposure pyramid attention-multi-exposure fusion)。该方法通过有效结合不同曝光水平的图像, 解决高对比度、低光照以及色彩和亮度平衡等问题。该方法由 3 个模块组成, 对 DPE (deep pyramid exposure) 进行改进。第 1 个模块专注于内容细节提取, 第 2 个模块负责色彩映射和校正, 第 3 个模块利用多尺度特征金字塔进行图像恢复。实验结果表明, 在不同的光照条件和动态场景下, DPEPA-MEF 能够更有效地融合多张曝光图像, 生成的图像在细节保留、色彩还原和对比度等方面表现出色。通过定量评估指标和主观视觉评估, DPEPA-MEF 均显示出显著的优势, 证明了该方法在实际应用中的巨大潜力和优越性。

关键词: 图像融合; 注意力机制; 金字塔网络; 多曝光; 细节提取; 色彩校正; 多尺度; 深度学习

中图分类号: TP181 **文献标志码:** A **文章编号:** 1673-4785(2025)04-0916-12

中文引用格式: 王卫苹, 刁亚鹏. 融合金字塔和多尺度注意力的多曝光图像融合优化算法 [J]. 智能系统学报, 2025, 20(4): 916-927.

英文引用格式: WANG Weiping, DIAO Yapeng. Optimized multi-exposure image fusion algorithm integrating pyramid and multi-scale attention[J]. CAAI transactions on intelligent systems, 2025, 20(4): 916-927.

Optimized multi-exposure image fusion algorithm integrating pyramid and multi-scale attention

WANG Weiping, DIAO Yapeng

(School of Computer and Communication Engineering, University of Science and Technology Beijing, Beijing 100083, China)

Abstract: The article introduces a multi-exposure image fusion technique named DPEPA-MEF (deep pyramid exposure pyramid attention-multi-exposure fusion). This technique aims to address image noise, blur, and detail loss caused by complex lighting conditions in real scenes. The DPEPA-MEF method effectively combines images with varying exposure levels to solve issues such as high contrast, low light, and color and brightness balance. It consists of three modules, which improve upon deep pyramid exposure (DPE). The first module focuses on content detail extraction, the second on color mapping and correction, and the third on image recovery using a multi-scale feature pyramid. Experimental results indicate that under different lighting conditions and dynamic scenes, DPEPA-MEF can more effectively fuse multiple exposure images. The resulting images exhibit excellent detail preservation, color reproduction, and contrast. Both quantitative evaluation metrics and subjective visual assessments demonstrate the significant advantages of DPEPA-MEF, confirming its great potential and superiority in practical applications.

Keywords: image fusion; attention mechanism; pyramid network; multi-exposure; detail extraction; color correction; multi-scale; deep learning

在现实场景中, 高对比度的情况普遍存在, 即明亮区域和暗区域同时出现。单张曝光的图像通常难以同时捕捉到所有细节, 导致高光区域可能过曝, 而阴影区域则可能欠曝, 进而导致细节丢失。尤其是在拍摄逆光场景时, 主体的过曝和背

景细节的缺失问题更加明显, 这限制了传统单曝光图像在复杂光照条件下的表现能力。

多曝光图像融合 (multi-exposure image fusion, MEF) 是一种有效且经济的图像质量增强技术。通过融合多张在不同曝光条件下拍摄的图像, MEF 技术能够充分利用亮区和暗区的信息, 显著扩展图像的动态范围。MEF 不仅可以整合不同曝光图像中的高光 and 阴影细节, 还能在复杂场景下有效保留前景和背景的细节信息。这一融合过程提升了细节的可见性和整体图像质量, 使得生

收稿日期: 2024-06-21. 网络出版日期: 2025-02-21.

基金项目: 工业和信息化部高质量发展专项 (TC220A04A-182); 国家自然科学基金项目 (62271045); 国家国际科技合作重点项目 (2022YFE0112300); 北京科技大学青年教师学科交叉研究项目 (FRF-IDRY-23-027); CCF-绿盟科技“鲲鹏”科研基金项目 (CCF-NS-FOCUS 2024017).

通信作者: 王卫苹. E-mail: weipingwangjt@ustb.edu.cn.

成的合成图像在曝光效果上更加均匀和平衡。

因此,研究和改进MEF技术,特别是在应对高对比度和复杂光照条件下的图像融合问题方面,有着重要的现实意义和研究价值。MEF技术大致分为传统方法和深度学习方法。

现有的传统方法往往依赖于手动提取的特征来执行融合过程,这使其在处理复杂和多变场景时显得力不从心。近年来,深度学习的进步促使了卷积神经网络(convolutional neural networks, CNNs)和生成对抗网络(generative adversarial networks, GANs)等技术在该领域的应用。这些基于深度学习的方法通过大量数据训练学习图像融合所需的映射关系。此类模型能够自动识别图像融合所需的特征和规则,从而提高融合过程的质量和效率。

深度学习解决方案在应对多曝光图像融合的挑战中得到了越来越广泛的应用。例如,Prabhakar等^[1-3]开发的DeepFuse的无监督网络,使用MEF-SSIM度量进行评估^[4]。此外,Han等^[5]引入了一种DPE(deep pyramid exposure)-MEF网络,旨在增强多曝光图像融合中的深度感知能力,其包含捕捉输入图像详细内容的模块以及用于调整颜色映射和最终输出校正的模块。

在MEF领域,对图像进行对齐是校正拍摄场景中微小位移或移动的关键步骤^[6-7]。错位或失真会导致重叠区域不匹配,从而产生明显的伪影。在涉及运动的场景中,这一问题尤为严重,伪影更容易出现。

鉴于上述问题,本文提出了一种旨在解决多曝光图像融合挑战的深度神经网络,对Han等^[5]最初提出的方法进行改进,提出了一种金字塔网络,结合注意力机制以解决MEF中的伪影问题。该网络利用不同尺度的图像金字塔,通过多次对原始图像进行下采样或上采样创建,这种金字塔结构能够捕捉不同尺度的细微差别和特征。将注意力机制集成到金字塔网络中,可以动态调节金字塔中每一层的影响力和权重。通过学习到的注意力权重,网络能够选择性地强调不同层次的细节,从而更有效地融合多曝光图像。

本文的主要贡献总结如下:

1) 为缓解伪影问题,提出了金字塔网络。金字塔网络能够在不同金字塔层次建模丰富的上下文信息。伪影问题通常与图像的全局和局部上下文密切相关。通过使用金字塔网络描述多个层次的上下文信息,可以更深入地理解图像的结构和语义,从而更精确地解决伪影问题。

2) 为了更准确的特征选择,提出了注意力机

制。通过计算每个特征的注意力权重,模型可以自动学习对当前任务最重要的特征,并减少对不相关或噪声特征的依赖。这有助于提高模型的准确性和鲁棒性。

3) 通过广泛的实验验证了所提方法的有效性,展示了其相较于当前最先进方法的优势。此外,进一步确认了该策略在提高单张图像曝光质量方面的可行性。

1 相关知识

传统方法通常包括图像变换、活动水平(或信息含量)的测量以及融合策略的制定^[8]。这些方法主要分为两类:在空间域中操作的方法^[9-10]和在变换域中操作的方法^[11-12]。

基于空间域的图像融合技术的核心思想是通过在空间域(像素级)操作和组合多个图像,将它们合并为一张图像。以Lee等^[13]的研究为代表,该研究探讨了图像融合中的权重分布,并研究了不同权重分布对融合结果的影响。最终提出了一种权重计算方法,可以根据图像的特征和重要性确定每个像素的权重。

基于变换域技术的核心思想是利用变换域表示的优势。例如,Krishnamoorthi等^[14]讨论了离散小波变换、平稳小波变换和变换采样离散小波等不同算法的发展和研究。通过实验测量了若干一致性参数,以评估所提出的融合技术的性能,证明了所提出的图像融合方法的优越性。

深度学习通过多层神经网络,凭借其强大的特征学习能力,为图像融合提供了新的机遇和挑战。GANs的出现进一步提升了图像融合的质量,使得高质量融合图像的生成成为可能。例如,Deng等^[15]提出的CF-Net网络,结合了曝光融合和图像超分辨率,通过耦合反馈架构实现,该设计通过反馈回路迭代地改进融合和超分辨率的结果。

图像融合算法的质量与图像融合的质量密切相关,包括源图像信息的保留程度、是否引入噪声和图像的清晰度^[16-17]。融合图像的评估可以分为主观评估和客观评估两种方法^[18-20]。

主观评估(或定性评估)通过人类视觉感知来判断图像质量,在多焦点图像融合研究中起着关键作用。然而,这种方法耗时且无法自动批量评估图像,同时由于观察者标准的个体差异,容易产生评估偏差。因此,在实际的多焦点图像融合评估中,通常结合使用主观评估和客观评估。

客观评估(或定量评估)使用特定的指标来衡量多焦点图像融合算法的性能。各种指标从不同角度评估算法:信息论、图像特征、图像结构相似

性和人类感知。本文采用峰值信噪比 (peak signal-to-noise ratio, PSNR)——基于信息论和结构相似性指数 (structure similarity index measure, SSIM)——基于图像特征进行定量分析,同时结合一些无监督的图像评估方法。

1) 改进的结构相似度 (S_a)

在基于文献 [21-23] 所提出的结构相似度的基础上,文献 [24-25] 中将其针对图像融合评价进行了改进。初始的结构相似度用来衡量图像间的相似性,包含亮度、对比度和结构 3 个部分。两张图像 (X, Y) 间结构相似度计算公式为

$$S(X, Y) = \frac{\mu_x \mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \cdot \frac{\sigma_x \sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \cdot \frac{\sigma_{xy} + C_3}{\sigma_x \sigma_y + C_3}$$

式中: μ_x 和 μ_y 分别表示图像 X 和 Y 的所有像素值平均值, σ_x 和 σ_y 分别表示图像 X 和 Y 的所有像素值标准差, σ_{xy} 表示两张图像间的协方差, C_1 、 C_2 和 C_3 是 3 个常数。式中 3 个部分分别衡量图像间的亮度、对比度和结构信息。

文献 [26-28] 提出的融合图像 (F) 和两张源图像 (I_1, I_2) 的结构相似度计算公式为

$$S_a(I_1, I_2, F) = \frac{S(I_1, F) + S(I_2, F)}{2}$$

式中 S_a 的值越接近 1, 则说明融合图像和源图像的结构越相近。

2) PSNR

PSNR 表示的是融合后的图像中, 峰值功率和噪声功率的比值, 可以用来测量多对焦图像融合过程中的失真程度, 具体定义为

$$P = 10 \lg \frac{r^2}{M_{SE}}$$

式中: r 是融合后图像中像素的最大值, 假如一张图片由 n 位来表示, 那么则有 $r = 2^n - 1$; M_{SE} 是均方差, 定义为

$$M_{SE} = \frac{M_{SE}(A, F) + M_{SE}(B, F)}{2}$$

A 和 B 表示两张原始图像, 此外:

$$M_{SE}(X, F) = \frac{1}{MN} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} (X(i, j) - F(i, j))^2$$

式中: M 和 N 分别表示图像的高度和宽度, $X(i, j)$ 和 $F(i, j)$ 分别表示图像 X 和图像 F 在位置 (i, j) 处的像素值。

通常情况下, PSNR 的值越大, 表示融合后的图像与源图像越接近, 多对焦图像融合的失真较小; 相反, PSNR 值较小则表示失真较大。因此, PSNR 值越大, 意味着多对焦图像融合算法输出的图像性能越好。

2 DPEPA-MEF 算法原理

本节描述 DPEPA-MEF 的网络, 用于在无监督环境下提高融合图像的感知质量, 并消除在融合过程中产生的伪影。

图 1 给出了 DPEPA-MEF 网络结构。该网络既要具有信息性, 又要在视觉上美观。从功能角度来看, 本文逻辑上将整个网络分为 3 个子网, 即细节增强模块 (detail enhancement module, DEM)、颜色增强模块 (color enhancement module, CEM) 和金字塔注意力重建模块 (pyramid attention module, PAM)。根据 3 个关键问题, 将整个网络分为 3 个小模块, 在一定程度上降低了复杂性。接下来, 将详细介绍这 3 个模块。

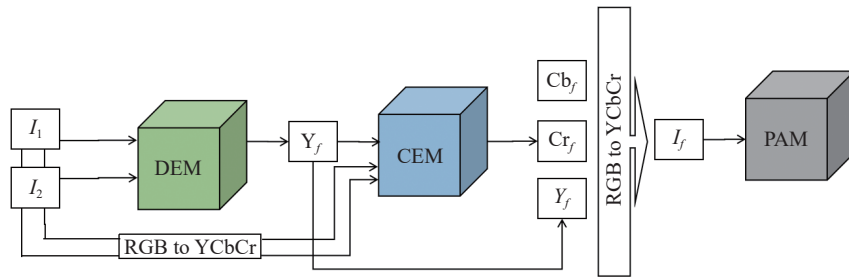


图 1 DPEPA-MEF 网络

Fig. 1 DPEPA-MEF network structure

2.1 细节增强模块

在 DEM 中, 针对曝光不足和曝光过度图像的细节问题提出了一种全新的处理方法。对于曝光不足的图像, 高动态范围的信息被压缩到有限的范围内; 而对于曝光过度的图像, 信息被上移并部分截断, 导致对比度低和细节损失。由于在 MEF 任务中没有真实的图像可用作优化目标, 核心任

务是找到一种方法来充分挖掘源图像中的信息, 增强对比度并保留细节, 以指导 DEM 网络的优化。

具体而言, 首先通过对源图像进行反转处理。在反转后的图像中, 原本曝光过度的区域将显示为曝光不足。然后, 对反转后的图像应用相同的增强规则。通过计算相应的调整图, 可以得到向下增强版本。这种方法通过放大细节, 能够

有效处理曝光过度的区域。此外,本文采用了双向细节增强技术。这意味着每个源图像将具有两个增强的参考图像, R 和 R' 。通过向上增强,曝光不足区域的细节得到了显著扩展;而通过向下处理,曝光过度区域的细节得到了有效放大。具体细节增强规则为:

给定一幅图像 I , 通过 $I_\alpha = \alpha \cdot I$ 获得全局增益, 其中曝光调整比例为 α 。当 $\alpha > 1$ 时, 图像被提升到更高的曝光水平(更亮); 而当 $\alpha < 1$ 时, 图像被降低到更低的曝光水平(更暗)。需要注意的是, 在同一场景中的不同区域的照明可能会有很大的差异。在同一张图像中, 可能同时存在曝光过度、正常曝光和曝光不足的区域。通过全局调整 $\alpha > 1$, 虽然曝光不足的区域会被照亮, 但原本正常曝光的部分会由于数字图像的表达限制而变得曝光过度。换句话说, 在这种情况下, 很难找到一个最优的 α 。因此, 需要一个局部自适应规则。

受到 Retinex 理论的启发, 一幅图像可以分解为两层, 即反射率和照度。本文将图像 I 分解为 $I = R \circ E$, 其中 R 和 E 分别表示场景细节和曝光分量, 运算符 \circ 表示逐元素乘积。通过简单的代数变换, 得出 $R = \frac{1}{E} \circ I$, 其中 $\frac{1}{E}$ 是 E 的逐元素反转。为便于解释用 A 表示 $\frac{1}{E}$ 。由于 A (或等效的 E) 是空间变化的, 因此调整也是空间变化的。注意, 如果 A 中的所有元素都具有相同的值, 则调整退化为全局调整。

2.2 颜色增强模块

在图像处理领域, 颜色增强模块在图像质量的主观评价中发挥着举足轻重的作用。与红外和可见光融合等其他图像融合任务不同, MEF 的目标是产生视觉上吸引人的结果。因此, 确保融合图像呈现生动真实的色彩对于增强 MEF 算法的视觉性能至关重要。然而, 许多现有的 MEF 方法往往忽视了这一方面。

传统上, 源图像被转换到 YCbCr 颜色空间, 融合技术主要应用于亮度 (Y) 通道, 而色度通道 (Cb 和 Cr) 以相对简单的方式处理, 通常使用加权平均。然而, 当源图像的曝光质量较差时, 由于数字设备的限制, 颜色信息可能会受到影响, 并且不同照明条件下的颜色可能会有所不同。在这些情况下, 直接通过加权平均得出的颜色通常不是最佳的。

为了解决这个问题, 开发了 CEM。该模块旨在通过学习融合图像亮度和源图像之间的颜色关系来找到最合适的颜色设置。CEM 检查来自两个源图像 (涵盖亮度和色度分量) 和目标亮度的所有数据, 以确定最佳颜色输出, 具体表达为

$$[Cb_f, Cr_f] = N_{CEM}(I_1, I_2, Y_f, \theta_{CEM})$$

式中: N_{CEM} 表示 CEM 网络, 参数 θ_{CEM} 表示 CEM 网络的参数。 Cb_f 和 Cr_f 代表色度数据, Y 代表亮度数据, 将 CEM 设置为一个每层 4 层的联合编码器-解码器结构, 以探索输入图像之间的颜色映射关系。

在实践中, 由于没有融合图像的标准颜色真值, CEM 无法通过直接训练来完成。但可以利用由相机拍摄的同一场景的多张不同曝光的图像, 这些图像提供了相对准确的颜色信息。为训练 CEM 从一个序列中随机选择 3 张图像, 输入其中的两张及第 3 张图像的亮度, 第 3 张图像的色度分量作为参考。通过这种方法, CEM 可以通过最小化预测值和参考值之间的差异进行训练, 使用曼哈顿距离作为度量标准。差异可以定义为

$$l_{CEM} = \|Cb_f - Cb_3\|_1 + \|Cr_f - Cr_3\|_1$$

可以看出, CEM 的训练可以免于参数调整。当然, 也可以调整两项之间的权重, 但考虑到 Cb (相对于绿的蓝) 和 Cr (相对于绿的红) 的定义, 我们倾向于平等对待它们。

在实际训练中, 训练图像被调整为 512×512 大小, 批量大小设置为 32, 补丁大小为 128×128 , 并进行数据增强 (随机翻转、旋转、调整大小和裁剪) 使用 ADAM 优化器, 并设定默认参数和固定学习率 1×10^4 , 窗口半径 r 设为 5, 并均匀使用预定义集, 通过改变参数得到不同的曝光增强比。

通过上述方法, CEM 能够有效地从源图像中学习色彩信息, 并将其应用于融合图像中。这样, 融合后的图像不仅在亮度上有所改善, 在色彩上也更加生动和真实, 显著提升了视觉感受。

2.3 金字塔注意力重建模块

PAM 作为一种简单而通用的构建块, 用于详尽地捕获远程依赖关系, 该方法充分利用了传统非局部操作的优点, 但更符合图像恢复的本质。具体来说, 原始的特征映射扩展到多尺度的特征金字塔。提出的操作通过搜索整个金字塔, 详尽地评估多个指定尺度上特征之间的相关性。

在金字塔注意力 (pyramid attention, PA) 重建模块中, 非局部注意力通过平均整个图像的特征来计算响应。形式上, 给定输入特征图 x , 该操作定义为

$$y^i = \frac{1}{\sigma(x)} \sum_j \phi(x^i, x^j) \theta(x^j)$$

式中 i 和 j 分别是输入 x 和输出 y 的索引。函数 ϕ 计算两个输入特征之间的成对相似度。 θ 是一个特征变换函数, 用于生成新的 x^j 表示。输出响

应 y^j 显式地对所有位置求和来获取所有特征的信息,并由标量函数 $\sigma(x)$ 进行归一化。尽管上述操作能够捕捉远程关联,但信息仅在单一尺度上提取。因此,它无法利用不同空间大小的多个信息区域之间的关系。

为了打破这种尺度限制,本实例基于 PA 提出的 PANet,如图 2 所示,PA 通过连续计算 S-A(scale agnostic)注意力来捕捉多尺度相关性,它

捕捉了不同尺度之间的相关性。在金字塔注意力中,计算目标特征和区域之间的亲和力。因此,响应特征是输入映射内多尺度对应的加权和。形式上,给定一系列比例因子 $S = \{1, s_1, s_2, \dots, s_n\}$,金字塔注意力可表示为

$$y^j = \frac{1}{\sigma(x)} \sum_{s \in S} \sum_j \phi(x^i, x_{\delta(s)}^j) \theta(x_{\delta(s)}^j)$$

式中 $\delta(s)$ 表示以输入 x 的索引 j 为中心的 s_2 的邻域。

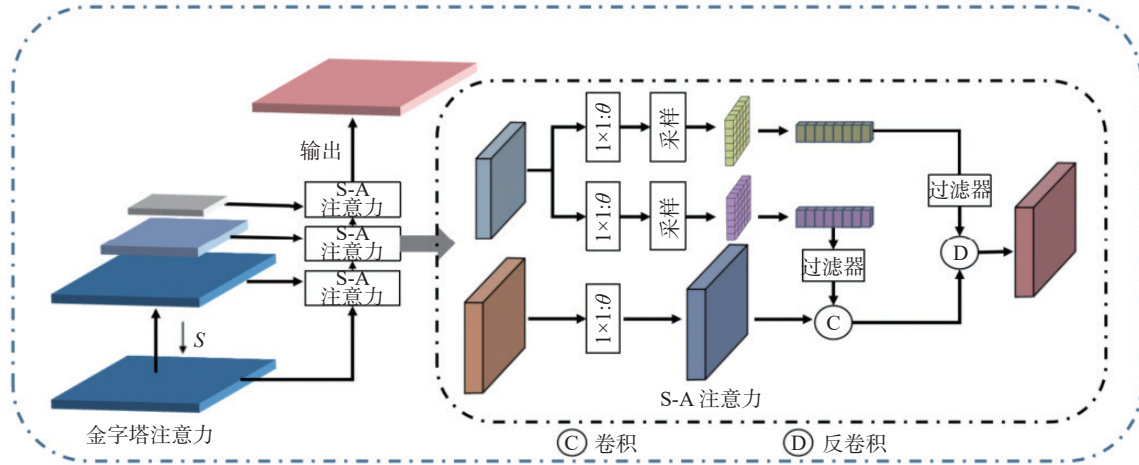


图 2 基于金字塔注意力提出的 PANet

Fig. 2 PANet with the proposed pyramid attention

换句话说,PA 以非局部多尺度的方式表现,通过显式处理规模金字塔 S 指定的所有位置 j 的更大区域。请注意,当仅指定单个尺度因子 $S = 1$ 时,提议的注意力降级为当前的非局部操作。因此,本实例提出的方法是一种更通用的操作,可以让网络充分享受自然图像的预测能力。

在 PA 模块中,我们通过 3×3 的小补丁在目标特征周围计算相关性,结合嵌入高斯函数来计算特征间的亲和力。同时,为了提高匹配过程的鲁棒性,我们采用了区域到区域的匹配方法,这种方法通过增加邻域相似性约束,使得匹配过程更加稳健,能够有效抑制噪声信号。该模块的计算过程完全可微,可以通过反向传播与网络一同优化。实验结果表明,即使在一个简单的 ResNet 骨干网络中仅插入一个 PA 模块, PANet 在多个图像修复任务上仍能显著优于当前最先进的方法,显示了该模块在提升图像恢复质量方面的强大优势。

2.4 损失函数

在其结构和损失函数中,DEM(差异增强模块)期望通过输入两个不同曝光源图像 I_1 和 I_2 , 生成一个具有更丰富细节的“良好”亮度分量 Y_f 用于融合图像,而色度分量则由 CEM(色度增强模块)负责。DEM 的功能可以表示为

$$Y_f = N_{\text{DEM}}(I_1, I_2, \theta_{\text{DEM}})$$

式中 N_{DEM} 表示以参数 a 作为学习参数的 DEM 网络。

本文采用类似于 UNet 的编码器-解码器结构作为核心架构^[29-31]。更具体地说,DEM 配备了两个编码器。其中一个被称为联合编码器,旨在同时处理两个源图像,以提取两幅图像中的相关特征;另一个称为判别编码器,对每个源图像分别进行编码,旨在从每个输入中找到代表性信息。然后,解码器将两个编码器的输出特征作为输入,并在每个尺度上接收编码器的跳跃连接,以生成最终融合图像的亮度分量。

为了指导 DEM 的训练(本部分考虑两次曝光融合)以产生期望的结果,损失函数通过双向增强规则考虑源图像的亮度以及每个源亮度的细节增强参考,总共涉及 6 个参数, $\hat{R}_q \in \{\hat{Y}_1, \hat{Y}_{1\text{inv}}, \hat{Y}_2, \hat{Y}_{2\text{inv}}, Y_1, Y_2\}$, 这可以从不同的角度规范学习。因此,损失函数可以表示为

$$L_{\text{DEM}} = \sum_{q=1}^Q \gamma_q (l_{\text{pix}}(\hat{R}_q, Y_f) + l_{\text{per}}^q(\hat{R}_q, Y_f))$$

式中 l_{pix} 表示 \hat{Y}_q 和 Y_f 之间的归一化曼哈顿距离:

$$l_{\text{pix}}(\hat{R}_q, Y_f) = \frac{1}{HW} \|\hat{R}_q - Y_f\|_1$$

式中: H 和 W 分别表示输入图像的高度和宽度,与输出图像相同; l_{per}^q 代表感知损失,定义为

$$l_{\text{per}}^q(\hat{R}_q, Y_f) = \sum_i \frac{1}{C_i H_i W_i} \|\phi_l(Y_f) - \phi_l(\hat{R}_q)\|_1 \quad (1)$$

式中: ϕ_l 为感知网络的第 l 层, C_l 、 H_l 、 W_l 为第 l 层张量特征图的维度。本文采用预训练的 VGG-19 网络进行感知特征提取, 其中 l 表示 {conv1_1, conv2_1, conv3_1, conv4_1, conv5_1} 层索引, 从式(1)中可以看出, 它同时考虑了深度特征域和原始图像域的引导。

3 实验验证与分析

3.1 实验设置

所提出的 DPE-MEF 在 SICE 数据集上进行了训练和评估^[32-34], 该数据集提供了 589 个室内和室外场景的多曝光图像序列。每个序列中的图像均由消费级相机拍摄, 并且对齐良好。

为训练 DEM 和 CEM, 从中随机选择 489 个序列用于训练, 其余的 100 个序列用于测试。在每个序列中, 选择曝光差异较大的图像对作为测试集。这种选择是因为它们具有更高的挑战性, 并且能够更有效地评估 MEF 算法在提取细节和保持全局结构方面的能力。框架在 PyTorch 中实现, 使用 Nvidia 3060 GPU 进行基于学习的竞争者的测试和训练。由于逻辑分区, DEM 和 CEM 是

分别训练的。

在训练 PAM 时, 所选择的数据集是 COCO unlabeled2017。这个数据集是 COCO (Common Objects in Context) 数据集中的未标注图像子集。COCO 数据集是一个广泛使用的计算机视觉数据集, 包含大量图像及其对应的注释信息, 被用于目标检测、图像分割、图像生成等任务。然而, 在 COCO 数据集中, 并非所有图像都标注有目标边界框或像素级分割掩码。这些未标注的图像被归类为 COCO unlabeled2017 数据集, 通常用于无监督学习、自监督学习、弱监督学习等任务。COCO unlabeled2017 数据集中的图像没有对应的注释信息, 因此在使用该数据集时, 通常需要依赖自动化方法或其他无监督学习技术进行图像分析和处理。这些未标注的图像可以用于训练无监督学习算法, 以探索图像中的特征、结构或模式, 或者生成新的数据样本。

3.2 定性比较

除了客观评价, 主观评价也是必要的。为了直观地展示本论文提出的 DPEPA 的优越性, 本节实验测试了 CF-NET、DPE 和 DPEPA 的输出效果, 如图 3 所示。



图 3 输出效果

Fig. 3 Output effect

从融合结果可以看出, CF-Net 的效果最差, 特别是与热气球上的文字相比, 可以发现伪影问题最为严重。由于网络结构设计过于简单, 难以处理具有运动的多焦点图像融合问题, 导致最终融合结果出现混乱, 不仅色彩失真, 而且模糊程度非常大, 导致输出的融合图像失去了最基本的观赏性。

图 4~7 给出了 4 个对比较为明显的测试样例。在图 4 的对比中, 除了 DPEPA 方法外, 其余 4 种方法在处理过曝光和欠曝光图像时表现出明

显的不足之处: CF-Net 方法在处理过曝光图像时, 部分细节丢失明显, 尤其是在高亮区域, 细节的保留不充分, 此外, 图像整体色调偏冷, 可能导致部分色彩信息的失真, 无法准确表现场景的真实颜色; DSIFT 方法在处理欠曝光区域时存在较大的问题, 亮度提升效果有限, 导致暗部细节无法有效恢复, 同时, 该方法在过曝光区域的表现也不尽如人意, 容易出现细节模糊和色彩失真, 影响整体视觉效果; GFF 方法在处理过曝光和欠曝光图像时, 虽然在细节保留上有所改善, 但仍

存在色彩还原不准确的问题, 图像色调偏蓝, 使得整体效果偏离真实场景; DPE 方法处理的图像存在一定程度的模糊现象, 这种模糊主要表现在过曝光区域和细节丰富的部分, 具体来说, DPE 方法在融合图像时, 可能没有充分保留图像的细

节纹理信息, 导致一些区域的清晰度下降。特别是在高亮区域, 图像的边缘和细节显得不够锐利, 整体图像显得有些模糊。这种模糊感不仅影响了图像的视觉效果, 还使得图像在表现真实场景时缺乏足够的清晰度和对比度。

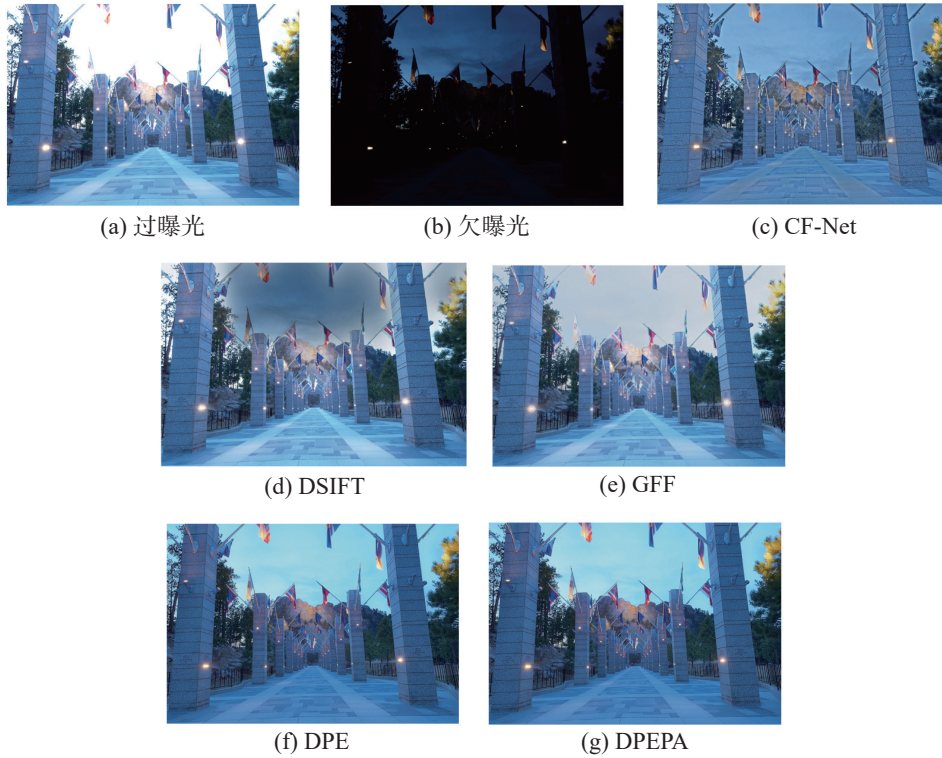


图 4 样例 1 曝光融合结果的视觉比较

Fig. 4 Visual comparison of exposure fusion results for sample one



图 5 样例 2 曝光融合结果的视觉比较

Fig. 5 Visual comparison of exposure fusion results for sample two



图 6 样例 3 曝光融合结果的视觉比较

Fig. 6 Visual comparison of exposure fusion results for sample three

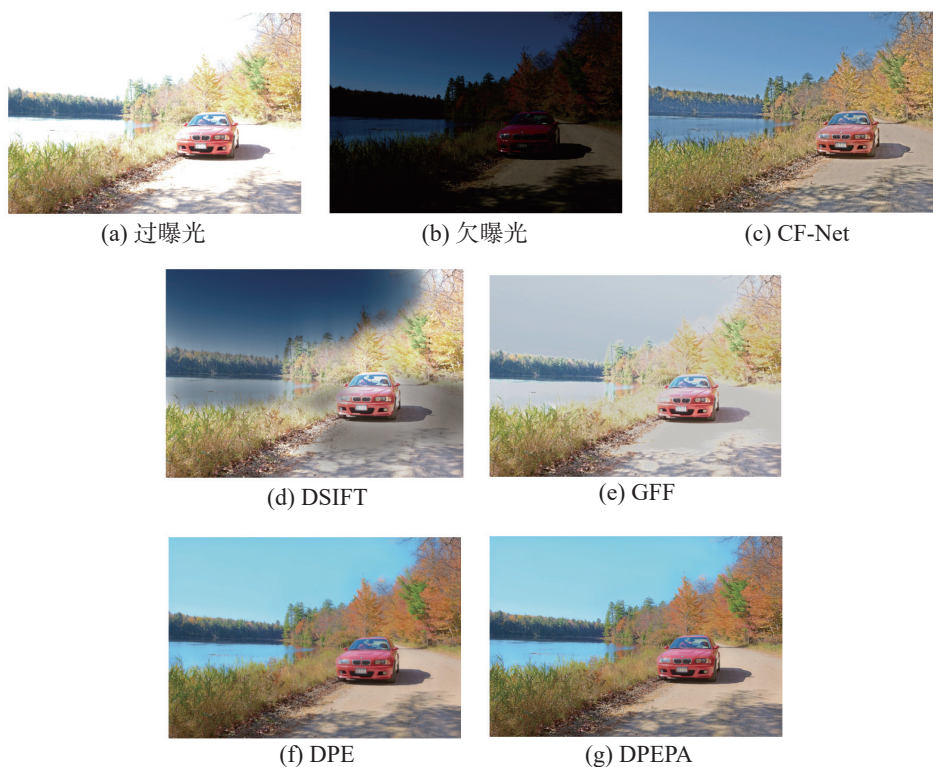


图 7 样例 4 曝光融合结果的视觉比较

Fig. 7 Visual comparison of exposure fusion results for sample four

如图 8 所示, 通过对 DPE 方法和 DPEPA 方法的局部放大图进行对比, 我们可以更直观地观察到 DPE 方法存在的模糊问题。特别是在车牌区域, DPE 方法生成的图像中, 车牌上的数字无

法清晰辨认, 整体显得模糊。这说明 DPE 方法在细节保留方面存在不足, 导致图像中的精细信息无法准确呈现。而在 DPEPA 方法生成的图像中, 车牌上的数字非常清晰, 可以轻松辨认。这显示

了 DPEPA 方法在细节处理和图像清晰度上的优势,使得生成的图像不仅视觉效果更好,也更具信息性和实用性。这种清晰度的差异充分说明了 DPEPA 方法在处理图像模糊问题上的优越性。



(a) DPE 局部放大图 (b) DPEPA 局部放大图

图 8 DPE 与 DPEPA 局部放大图对比

Fig. 8 Comparison of DPE and DPEPA localized enlarged image

相比之下, DPEPA 方法在处理过曝光和欠曝光问题上表现优异,不仅能够有效恢复图像中的细节,还能保持色彩的自然和准确,使得最终生成的图像在视觉上更加逼真和令人满意。

3.3 定量比较

如图 9、10 分别给出了样例 4、6 的直方图。在 DPE 方法的处理下,直方图中可以观察到显著的双峰分布,这表明在高亮度和低亮度区域内像素值的集中程度较高。这种尖锐的峰值集中现象特别明显,这可能是由于 DPE 在增强对比度时过

度强调了这些区域,导致细节的丢失或者伪影的出现。高亮度区域的峰值过于集中使得图像中的亮部细节消失或模糊,而阴影部分的细节也可能因为对比度的增加而受到影响。这种处理方式虽然能够显著增强图像的整体视觉冲击力,但在保留细节和真实感方面有所欠缺。相比之下, DPEPA 则通过更平滑的过渡处理,有效保留图像的细节和层次感,使亮度过渡更加自然。CF-Net 的处理强调在细节保留方面的能力,尤其在亮度过渡区间上的表现优于 DPE。然而,这种细致的处理方式可能会导致图像对比度不足,使整体效果不够鲜明,欠缺视觉上的冲击力。尽管如此, CF-Net 的细节保留能力在某些应用场景中仍然非常重要,尤其是在需要准确呈现物体形态和纹理的场合。DSIFT 方法在直方图中表现出不连续的峰值,这表明在某些亮度区间内,像素分布不均匀,可能由于算法在这些区间内进行了过度的处理或缺乏平滑性,使图像在视觉上不够自然。而 GFF 方法则主要集中在高亮度区间的增强,虽然图像更加明亮,但高亮区域容易过曝,细节可能丢失。DPEPA 则在这些方面取得了平衡,不仅增强了图像的对比度,同时避免了细节丢失,使图像在视觉效果和真实感之间取得了良好的平衡。

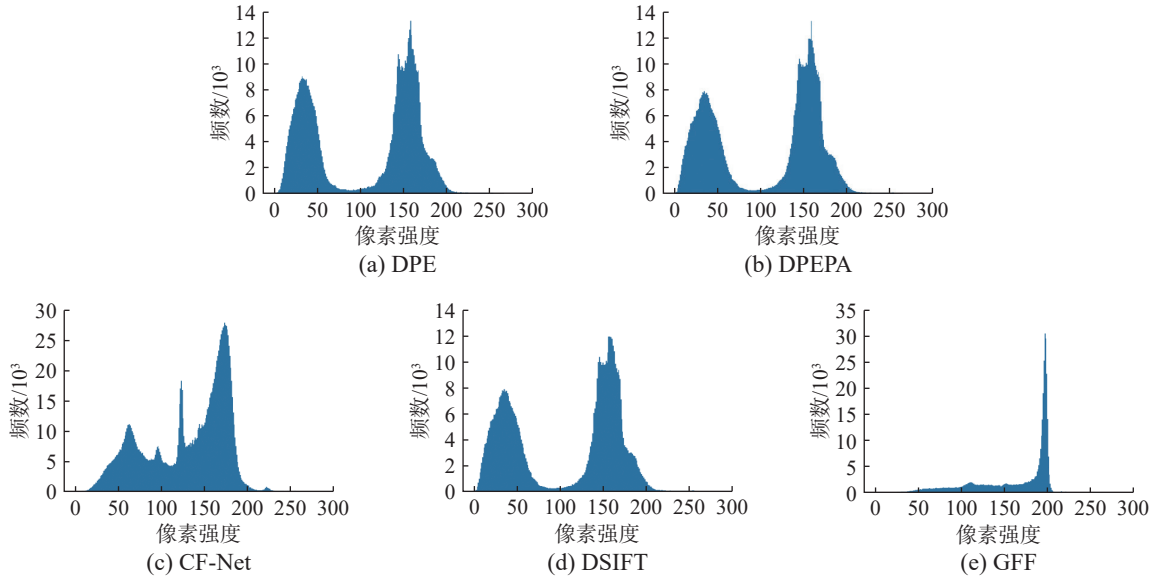
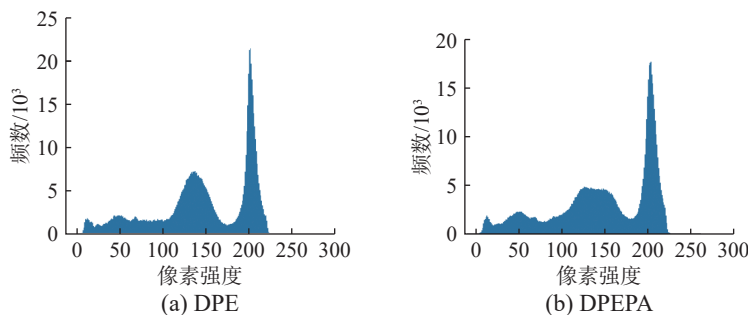


图 9 样例 4 像素强度直方图对比

Fig. 9 Comparison of sample four pixel intensity histogram



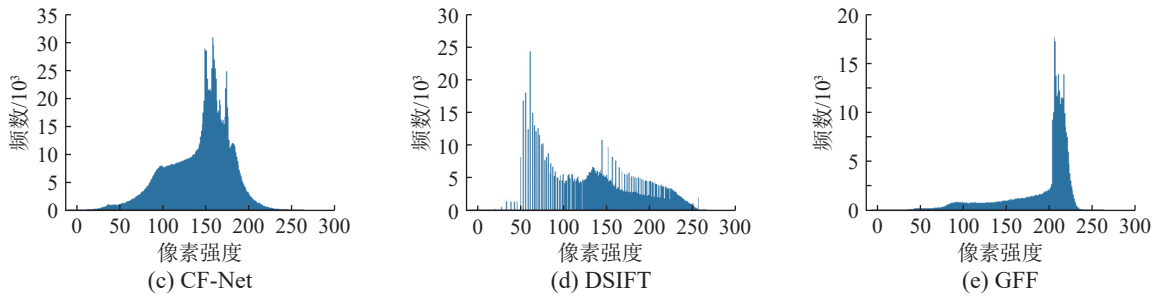


图 10 样例 6 像素强度直方图对比

Fig. 10 Comparison of sample six pixel intensity histogram

表 1 给出了 5 种不同的方法 (CF-Net、DSIFT、GFF、DPE 以及本文方法 DPEPA) 的对比结果。实验选择了 PSNR 和 SSIM 两个主要的定量指标。定量实验结果表明, 无论是 PSNR 还是 SSIM, 本文方法 DPEPA 方法均表现出最佳性能。

表 1 不同方法在 10 组样例上的指标比较

Table 1 Comparison of indicators of different methods on ten groups of samples

指标	样例	CF-Net	DSIFT	GFF	DPE	DPEPA
PSNR	样例1	18.61	13.49	17.24	33.14	33.24
	样例2	14.63	11.22	7.55	31.26	31.29
	样例3	22.39	14.90	7.85	34.83	34.91
	样例4	22.45	11.64	8.21	34.96	35.18
	样例5	19.59	11.22	9.64	33.87	33.99
	样例6	17.49	9.12	9.82	32.69	32.82
	样例7	18.51	9.10	8.49	32.09	32.09
	样例8	15.56	8.42	10.85	31.64	31.71
	样例9	17.67	12.18	6.51	33.22	33.25
	样例10	24.55	19.11	5.82	34.99	35.06
SSIM	样例1	0.8425	0.4621	0.7619	0.8185	0.9092
	样例2	0.8825	0.6304	0.3644	0.8536	0.9012
	样例3	0.9041	0.5478	0.2506	0.8225	0.8741
	样例4	0.7853	0.3361	0.2503	0.7929	0.8892
	样例5	0.8611	0.5851	0.2623	0.7641	0.8528
	样例6	0.8522	0.4321	0.3331	0.7835	0.8698
	样例7	0.8925	0.5279	0.4207	0.8086	0.8187
	样例8	0.7753	0.4652	0.3836	0.8061	0.8781
	样例9	0.8290	0.7298	0.3411	0.8796	0.8978
	样例10	0.8085	0.6867	0.2597	0.8628	0.8799

DPEPA 在所有对比方法中表现优越, 特别是在处理包含运动的多焦点图像融合任务中。相比之下, CF-Net、DSIFT、GFF 和 DPE 在多个样例中的 PSNR 和 SSIM 指标均不如 DPEPA。特别是 DPE-Net, 由于其主要适用于静态场景且依赖源图像的精确对齐, 因此在包含运动的场景中表现不佳。这导致了 DPE-Net 在提取运动信息时的能力不足, 无法达到 DPEPA 的效果。

作为经典的多曝光图像融合网络, DPE-Net 在本文中的 PSNR 和 SSIM 指标均低于 DPEPA。这可能是由于现有的 DPE-MEF 主要适用于静态场景, 需要源图像精确对齐。然而, 由于相机和物体的移动, 曝光序列中有时会出现前景和背景错位的情况, 这可能导致静态融合技术无法产生理想的融合效果。DPE-Net 在提取运动信息的能力上较弱, 因此效果不如本文中的 DPEPA。

4 结束语

本文提出了一种新的多曝光图像融合算法 DPEPA, 旨在解决图像伪影和复杂光照条件下的融合挑战。通过引入深度学习和注意力机制, 结合不同曝光水平的图像, 在高对比度和低光照场景中增强细节表现和平衡整体亮度。特别是, DPEPA 在处理多焦点图像时, 通过动态提取和综合不同曝光图像中的信息, 显著减少了图像伪影, 并提升了细节保留的准确性和视觉质量。大量实验结果证明, DPEPA 在生成高质量融合图像方面显著优于现有技术, 展现了在多种复杂场景下的广泛适应性和卓越性能。该方法的提出为多曝光图像融合领域提供了新的思路和解决方案, 具有重要的理论和实际应用价值。

参考文献:

- [1] YI Xunpeng, TANG Linfeng, ZHANG Hao, et al. Diff-IF: Multi-modality image fusion via diffusion model with fusion knowledge prior[J]. *Information fusion*, 2024, 110: 102450.
- [2] ZHANG Xingchen. Benchmarking and comparing multi-exposure image fusion algorithms[J]. *Information fusion*, 2021, 74: 111–131.
- [3] LEI Dajiang, RAN Gangsheng, ZHANG Liping, et al. A spatiotemporal fusion method based on multiscale feature extraction and spatial channel attention mechanism[J]. *Remote sensing*, 2022, 14(3): 461.
- [4] LIAN Zuozheng, WANG Haizhen. An image deblurring method using improved U-Net model based on multilayer fusion and attention mechanism[J]. *Scientific reports*, 2023, 13: 21402.
- [5] HAN Dong, LI Liang, GUO Xiaojie, et al. Multi-exposure image fusion via deep perceptual enhancement[J]. *Information fusion*, 2022, 79: 248–262.
- [6] WANG Zhao, LI Feng, CONG Runmin, et al. Adaptive feature fusion network based on boosted attention mechanism for single image dehazing[J]. *Multimedia tools and applications*, 2022, 81(8): 11325–11339.
- [7] ZHANG Jianming, FENG Wenjun, YUAN Tingyu, et al. SCSTCF: spatial-channel selection and temporal regularized correlation filters for visual tracking[J]. *Applied soft computing*, 2022, 118: 108485.
- [8] CAI Jianrui, GU Shuhang, ZHANG Lei. Learning a deep single image contrast enhancer from multi-exposure images[J]. *IEEE transactions on image processing*, 2018, 27(4): 2049–2062.
- [9] LI Weiqiang, MU Jiatong, LIU Guizhong. Multiple object tracking with motion and appearance cues[C]//2019 IEEE/CVF International Conference on Computer Vision Workshop. Seoul: IEEE, 2019: 161–169.
- [10] NIU Wenjia, ZHANG Kaihao, LUO Wenhan, et al. Deep robust image deblurring via blur distilling and information comparison in latent space[J]. *Neurocomputing*, 2021, 466: 69–79.
- [11] LI Hui, MA Kede, YONG Hongwei, et al. Fast multi-scale structural patch decomposition for multi-exposure image fusion[J]. *IEEE transactions on image processing*, 2020, 29: 5805–5816.
- [12] SHAO Mingwen, LI Le, MENG Deyu, et al. Uncertainty guided multi-scale attention network for raindrop removal from a single image[J]. *IEEE transactions on image processing*, 2021, 30: 4828–4839.
- [13] LEE K H, CHEN Xi, HUA Gang, et al. Stacked cross attention for image-text matching[C]//Proceedings of the European conference on computer vision. New York: Springer, 2018: 201–216.
- [14] KRISHNAMOORTHY R, BAI A, SRINIVAS A. Medical image fusion based on transformation domain approaches[C]//IOP Conference Series: Materials Science and Engineering. [S. l.]: IOP Publishing, 2020, 981(4): 042082.
- [15] DENG Xin, ZHANG Yutong, XU Mai, et al. Deep coupled feedback network for joint exposure fusion and image super-resolution[J]. *IEEE transactions on image processing*, 2021, 30: 3098–3112.
- [16] ZHANG Weixia, MA Kede, ZHAI Guangtao, et al. Uncertainty-aware blind image quality assessment in the laboratory and wild[J]. *IEEE transactions on image processing*, 2021, 30: 3474–3486.
- [17] CHEN Tianhai, YANG Xichen, LI Nengxin, et al. Underwater image quality assessment method based on color space multi-feature fusion[J]. *Scientific reports*, 2023, 13: 16838.
- [18] PAN Zhaoqing, YUAN Feng, LEI Jianjun, et al. VCR-Net: visual compensation restoration network for No-reference image quality assessment[J]. *IEEE transactions on image processing*, 2022, 31: 1613–1627.
- [19] PATHAN R K, BISWAS M, YASMIN S, et al. Sign language recognition using the fusion of image and hand landmarks through multi-headed convolutional neural network[J]. *Scientific reports*, 2023, 13(1): 16975.
- [20] HOFFMANN M, BILLOT B, GREVE D N, et al. SynthMorph: learning contrast-invariant registration without acquired images[J]. *IEEE transactions on medical imaging*, 2022, 41(3): 543–558.
- [21] MAR C C, ZIN T T, TIN P, et al. Cow detection and

- tracking system utilizing multi-feature tracking algorithm [J]. *Scientific reports*, 2023, 13(1): 17423.
- [22] TANG Linfeng, ZHANG Hao, XU Han, et al. Rethinking the necessity of image fusion in high-level vision tasks: a practical infrared and visible image fusion network based on progressive semantic injection and scene fidelity[J]. *Information fusion*, 2023, 99: 101870.
- [23] 刘万军, 佟畅, 曲海成. 空洞卷积与注意力融合的对抗式图像阴影去除算法[J]. *智能系统学报*, 2021, 16(6): 1081–1089.
LIU Wanjun, TONG Chang, QU Haicheng. An antagonistic image shadow removal algorithm based on dilated convolution and attention mechanism[J]. *CAAI transactions on intelligent systems*, 2021, 16(6): 1081–1089.
- [24] ZHANG Hao, XU Han, TIAN Xin, et al. Image fusion meets deep learning: a survey and perspective[J]. *Information fusion*, 2021, 76: 323–336.
- [25] QIU Sen, ZHAO Hongkai, JIANG Nan, et al. Multi-sensor information fusion based on machine learning for real applications in human activity recognition: State-of-the-art and research challenges[J]. *Information fusion*, 2022, 80: 241–265.
- [26] 王建, 吴锡生. 基于改进的稀疏表示和 PCNN 的图像融合算法研究[J]. *智能系统学报*, 2019, 14(5): 922–928.
WANG Jian, WU Xisheng. Image fusion based on the improved sparse representation and PCNN[J]. *CAAI transactions on intelligent systems*, 2019, 14(5): 922–928.
- [27] 姜义, 吕荣镇, 刘明珠, 等. 基于生成对抗网络的人脸口罩图像合成[J]. *智能系统学报*, 2021, 16(6): 1073–1080.
JIANG Yi, LYU Rongzhen, LIU Mingzhu, et al. Masked face image synthesis based on a generative adversarial network[J]. *CAAI transactions on intelligent systems*, 2021, 16(6): 1073–1080.
- [28] 张心祎, 谭耀, 邢向磊. 基于物理先验的深度特征融合水下图像复原[J]. *智能系统学报*, 2023, 18(6): 1185–1196.
ZHANG Xinyi, TAN Yao, XING Xianglei. Deep feature fusion for underwater-image restoration based on physical priors[J]. *CAAI transactions on intelligent systems*, 2023, 18(6): 1185–1196.
- [29] LI Hui, WU Xiaojun, KITTLER J. RFN-Nest: an end-to-end residual fusion network for infrared and visible images[J]. *Information fusion*, 2021, 73: 72–86.
- [30] TANG Linfeng, YUAN Jiteng, MA Jiayi. Image fusion in the loop of high-level vision tasks: a semantic-aware real-time infrared and visible image fusion network[J]. *Information fusion*, 2022, 82: 28–42.
- [31] QUAN Xiongwen, HOU Guangyao, YIN Wenya, et al. A multi-modal and multi-stage fusion enhancement network for segmentation based on OCT and OCTA images [J]. *Information fusion*, 2025, 113: 102594.
- [32] HE Xiaoyu, WANG Yong, ZHAO Shuang, et al. Co-attention fusion network for multimodal skin cancer diagnosis[J]. *Pattern recognition*, 2023, 133: 108990.
- [33] LIU Yun, ZHANG Xiaoming, ZHANG Qianyun, et al. Dual self-attention with co-attention networks for visual question answering[J]. *Pattern recognition*, 2021, 117: 107956.
- [34] XIANG Fengtao, JIAN Zhang, LIANG Pan, et al. Robust image fusion with block sparse representation and online dictionary learning[J]. *IET image processing*, 2018, 12(3): 345–353.

作者简介:



王卫苹, 教授, 主要研究方向为类脑计算、忆阻神经网络、联想记忆感知模拟、复杂网络、网络安全和图像加密。主持和参与国家自然科学基金项目、国际联合基金重点项目等多项, 获得中国通信学会最佳论文奖、河南省科技进步二等奖等荣誉。发表学术论文 90 余篇, 拥有软著与专利 20 余项。E-mail: weipingwangjt@ustb.edu.cn。



刁亚鹏, 硕士研究生, 主要研究方向为类脑计算、脑机接口和图像处理领域。E-mail: ypdiao@xs.ustb.edu.cn。